

## SEQUENCE LISTING

### (1) GENERAL INFORMATION:

- (i) APPLICANT:
  - (A) NAME: von Schaewen, Antje Dr. rer. nat.
  - (B) STREET: Natruperstrasse 169a
  - (C) CITY: Osnabrueck
  - (E) COUNTRY: Germany
  - (F) POSTAL CODE (ZIP): D-49076
  - (G) TELEPHONE: +49-541-684029
- (ii) TITLE OF INVENTION: Plant gntI sequences and the use thereof for the production of plants having reduced or lacking N-acetyl glucosaminyl transferase I (GnTI) activity
- (iii) NUMBER OF SEQUENCES: 6
- (iv) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

### (2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1669 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: NO
- (iv) ANTI-SENSE: NO
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Solanum tuberosum
  - (B) STRAIN: Desiree
  - (D) DEVELOPMENTAL STAGE: Sink organ
  - (F) TISSUE TYPE: Mesophyll
  - (G) CELL TYPE: Leaf cells
- (vii) IMMEDIATE SOURCE:
  - (A) LIBRARY: Lambda ZAP II (Eco RI)
  - (B) CLONE: gntI-A1(K)

### (ix) FEATURE:

- (A) NAME/KEY: misc feature
- (B) LOCATION: 659..667
- (D) OTHER INFORMATION: /function= "Asn codon in this context is a potential glycosylation site"
- /product= "N-glycosylation consensus sequence"
- /phenotype= "N-glycans modulate protein properties"
- /standard\_name= "N-glycosylation site"
- /label= pot-CHO
- /note= "GnTI-coding sequences from animals do not contain this feature"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:53..1393
- (C) IDENTIFICATION METHOD: experimental
- (D) OTHER INFORMATION:/codon\_start= 53
  - /function= "initiates complex N-glycans on secretory glycoproteins"
  - /EC\_number= 2.4.1.101
  - /product= "beta-1,2-N-acetylglucosaminyltransferase I"
  - /evidence= EXPERIMENTAL
  - /gene= "cgl"
  - /standard\_name= "gntI"
  - /label= ORF
  - /note= "first gntI sequence from potato (unpublished)"

(ix) FEATURE:

- (A) NAME/KEY: 5'UTR
- (B) LOCATION:15..52

(ix) FEATURE:

- (A) NAME/KEY: 3'UTR
- (B) LOCATION:1394..1655

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:80..139
- (D) OTHER INFORMATION:/function= "membrane anchor (amino acids 10-29)"
  - /product= "hydrophobic amino acid stretch in GntI"
  - /standard\_name= "membrane anchor of a type II Golgi protein"
  - /note= "identified by comparison with GntI sequences from animals"

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION:1..14
- (D) OTHER INFORMATION:/function= "used for cloning the cDNA library in Lambda ZAPII"
  - /product= "EcoRI/NotI-cDNA adapter"
  - /number= 1

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION:1656..1669
- (D) OTHER INFORMATION:/product= "EcoRI/NotI-cDNA adapter"
  - /number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAATTCGCGG CCGCCTGAGA AACCTCGAA TTCAATTTTCG CATTGGCAG AG ATG  
Met  
1

55

AGA GGG AAC AAG TTT TGC TTT GAT TTA CGG TAC CTT CTC GTC GTG GCT  
Arg Gly Asn Lys Phe Cys Phe Asp Leu Arg Tyr Leu Leu Val Val Ala  
5 10 15

103

GCT CTC GCC TTC ATC TAC ATA CAG ATG CGG CTT TTC GCG ACA CAG TCA  
Ala Leu Ala Phe Ile Tyr Ile Gln Met Arg Leu Phe Ala Thr Gln Ser  
151

151

20 25 30 199  
GAA TAT GTA GAC CGC CTT GCT GCT GCA ATT GAA GCA GAA AAT CAT TGT  
Glu Tyr Val Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His Cys  
35 40 45  
247  
ACA AGT CAG ACC AGA TTG CTT ATT GAC AAG ATT AGC CAG CAG CAA GGA  
Thr Ser Gln Thr Arg Leu Leu Ile Asp Lys Ile Ser Gln Gln Gln Gly  
50 55 60 65  
295  
AGA GTA GTA GCT CTT GAA GAA CAA ATG AAG CAT CAG GAC CAG GAG TGC  
Arg Val Val Ala Leu Glu Glu Gln Met Lys His Gln Asp Gln Glu Cys  
70 75 80  
343  
CGG CAA TTA AGG GCT CTT GTT CAG GAT CTT GAA AGT AAG GGC ATA AAA  
Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile Lys  
85 90 95  
391  
AAG TTA ATC GGA GAT GTG CAG ATG CCA GTG GCA GCT GTA GTT GTT ATG  
Lys Leu Ile Gly Asp Val Gln Met Pro Val Ala Ala Val Val Val Met  
100 105 110  
439  
GCT TGC AGT CGT ACT GAC TAC CTG GAG AGG ACT ATT AAA TCC ATC TTA  
Ala Cys Ser Arg Thr Asp Tyr Leu Glu Arg Thr Ile Lys Ser Ile Leu  
115 120 125  
487  
AAA TAC CAA ACA TCT GTT GCA TCA AAA TAT CCT CTT TTC ATA TCC CAG  
Lys Tyr Gln Thr Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser Gln  
130 135 140 145  
535  
GAT GGA TCA AAT CCT GAT GTA AGA AAG CTT GCT TTG AGC TAT GGT CAG  
Asp Gly Ser Asn Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Gly Gln  
150 155 160  
583  
CTG ACG TAT ATG CAG CAC TTG GAT TAT GAA CCT GTG CAT ACT GAA AGA  
Leu Thr Tyr Met Gln His Leu Asp Tyr Glu Pro Val His Thr Glu Arg  
165 170 175  
631  
CCA GGG GAA CTG GTT GCA TAC TAC AAG ATT GCA CGT CAT TAC AAG TGG  
Pro Gly Glu Leu Val Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys Trp  
180 185 190  
679  
GCA TTG GAT CAG CTG TTT CAC AAG CAT AAT TTT AGC CGT GTT ATC ATA  
Ala Leu Asp Gln Leu Phe His Lys His Asn Phe Ser Arg Val Ile Ile  
195 200 205  
727  
CTA GAA GAT GAT ATG GAA ATT GCT GCT GAT TTT TTT GAC TAT TTT GAG  
Leu Glu Asp Asp Met Glu Ile Ala Ala Asp Phe Phe Asp Tyr Phe Glu  
210 215 220 225  
775  
GCT GGA GCT ACT CTT CTT GAC AGA GAC AAG TCG ATT ATG GCT ATT TCT  
Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile Ser  
230 235 240  
823  
TCT TGG AAT GAC AAT GGA CAA AGG CAG TTC GTC CAA GAT CCT GAT GCT  
Ser Trp Asn Asp Asn Gly Gln Arg Gln Phe Val Gln Asp Pro Asp Ala  
245 250 255

CTT TAC CGC TCA GAC TTT TTT CCT GGT CTT GGA TGG ATG CTT TCA AAA Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser Lys 260 270	871
TCA ACT TGG TCC GAA CTA TCT CCA AAG TGG CCA AAG GCT TAC TGG GAT Ser Thr Trp Ser Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp 275 285	919
GAC TGG CTA AGG CTG AAA GAA AAT CAC AGA GGT CGA CAA TTT ATT CGC Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile Arg 290 295 305	967
CCA GAA GTT TGC AGA ACG TAC AAT TTT GGT GAG CAT GGT TCT AGT TTG Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu 310 315 320	1015
GGG CAG TTT TTT AAG CAG TAT CTT GAG CCA ATT AAG CTA AAT GAT GTC Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp Val 325 330 335	1063
CAG GTT GAT TGG AAG TCA ATG GAC CTA AGT TAC CTT TTG GAG GAC AAC Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp Asn 340 345 350	1111
TAT GTG AAA CAC TTT GGC GAC TTG GTT AAA AAG GCT AAG CCC ATC CAC Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile His 355 360 365	1159
GGA GCT GAT GCT GTT TTG AAA GCA TTT AAC ATA GAT GGT GAT GTG CGT Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val Arg 370 375 380 385	1207
ATT CAG TAC AGA GAC CAA CTA GAC TTT GAA GAT ATC GCT CGA CAG TTT Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asp Ile Ala Arg Gln Phe 390 395 400	1255
GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGG GCA GCA TAT AAA Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr Lys 405 410 415	1303
GGG ATA GTA GTT TTC CGG TTT CAA ACA TCT AGA CGT GTG TTC CTT GTT Gly Ile Val Val Phe Arg Phe Gln Thr Ser Arg Arg Val Phe Leu Val 420 425 430	1351
TCC CCT GAT TCT CTT CGA CAA CTT GGA GTT GAA GAT ACT TAG Ser Pro Asp Ser Leu Arg Gln Leu Gly Val Glu Asp Thr * 435 440 445	1393
CGAAGATATG ATTGGAGCCT GAGCAACAAT TTAGACTTAT TTGGTAGGAT ACATTTGAAA	1453
GAGCTGACAC GAAAAGTATG ACTACCAGTA GCTACATGCA ACATTTTAAAT GTTAATGGAA	1513
GGAACCCACT GCTTATTGTT GGAATGGATG AATCATCACC ACATCCTATT ATTCAAGTTT	1573
ACAAACATAA AGAGGAAATG TTGCCCTATA AAAACAAATT TTTTGTTCCT AAGAAGGAAC	1633
GTTACGATTA TGAGCAACTT TGGCGGCCGC GAATTG	1669

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 447 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

(11) SEQUENCE DESCRIPTION: SEQ

Met Arg Gly Asn Lys Phe Cys Phe Asp Leu Arg Tyr Leu Leu Val 15  
1 5

Ala Ala Leu Ala Phe Ile Tyr Ile Gln Met Arg Leu Phe Ala Thr Gln 30  
20 25

Ser Glu Tyr Val Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu Asn His 45  
35 40

Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Lys Ile Ser Gln Gln Gln 60  
50 55

Gly Arg Val Val Ala Leu Glu Glu Gln Met Lys His Gln Asp Gln Glu 80  
65 70 75

Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile 95  
85 90

Lys Lys Leu Ile Gly Asp Val Gln Met Pro Val Ala Ala Val Val Val 110  
100 105

Met Ala Cys Ser Arg Thr Asp Tyr Leu Glu Arg Thr Ile Lys Ser Ile 125  
115 120

Leu Lys Tyr Gln Thr Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser 140  
130 135

Gln Asp Gly Ser Asn Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Gly 160  
145 150 155

Gln Leu Thr Tyr Met Gln His Leu Asp Tyr Glu Pro Val His Thr Glu 175  
165 170

Arg Pro Gly Glu Leu Val Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys 190  
180 185

Trp Ala Leu Asp Gln Leu Phe His Lys His Asn Phe Ser Arg Val Ile 205  
195 200

Ile Leu Glu Asp Asp Met Glu Ile Ala Ala Asp Phe Phe Asp Tyr Phe 220  
210 215

Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile 240  
225 230 235

Ser Ser Trp Asn Asp Asn Gly Gln Arg Gln Phe Val Gln Asp Pro Asp 255  
245 250

Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser 270  
260 265

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:  
(b) LENGTH: 1737 base pa

- SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1737 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
ORGANISM: N

- ORIGINAL SOURCE:  
(A) ORGANISM: *Nicotiana tabacum*  
(B) STRAIN: Samsun NN  
(D) DEVELOPMENTAL STAGE: Sink organ  
(F) TISSUE TYPE: Mesophyll  
(G) CELL TYPE: Leaf cells

(vii) IMMEDIATE SOURCE:  
LIBRARY: Lam

- IMMEDIATE SOURCE:  
(A) LIBRARY: Lambda ZAP II (Eco RI)  
(B) CLONE: gntI-A9(T)

URE:  
(A) NAME/KEY: misc feature  
733 741

(D) OTHER INFORMATION: /function= "Asn codon  
is a potential glycosylation site"

```
/product= "N-glycosylation consensus"
/feature_type= "N-glycans modulate protein"
```

```

/phenotype= "N-glycans"
properties"
name= "N-glycosylation site"

```

```

/phenotype= "N-glycosylation site"
properties"
/standard name= "N-glycosylation site"

```

```
/standard_name=  
/label= pot-CHO
```

```

/standard_name= "N-glycosylation"
/label= pot-CHO
/note= "GnTI sequences from animals do not contain this
feature"

```

FEATURE:  
(A) NAME/KEY: CDS

(A) NAME/REF: ...  
(B) LOCATION: 127...1467  
IDENTIFICATION METHOD: experimental  
start= 127

(C) IDENTIFICATION METHOD: experimental  
ADDITIONAL INFORMATION: /codon\_start= 127

(D) OTHER INFORMATION: /codon\_start= 127  
function= "initiates complex N-glycans on

```

/function= "initiates co
secretory glycoproteins"

```

secretory glycoprotein  
/EC number= 2.4.1.101

```
/EC_number=
/product=
```

```

/EC_number= 2.4.1.10
/product=
1,2-N-acetylglucosaminyltransferase I"

```

"beta-1,2-N-acetylglucosaminidase  
/evidence= EXPERIMENTAL

```
/evidence=
/gene= "cgl"
```

```
/gene= "cglI"
/standard name= "gntI"
```

```
/label= ORF
```

```

/standard_name="gntI"
/label= ORF
/note= "first gntI sequence from tobacco (unpublished)"

```

FEATURE:  
(A) NAME/KEY: 5'UTR  
REGION: 15 12

(A) NAME/KEY: 3 GIR  
(B) LOCATION: 15..126

FEATURE:  
(A) NAME/KEY: 3'UTR

(A) NAME/KEY: 301R  
(B) LOCATION: 1468..1723

FEATURE:  
(A) NAME/KEY: CDS

(A) NAME/KEY: CDS  
(B) LOCATION: 154..213  
ADDITIONAL INFORMATION

(D) OTHER INFORMATION: /function= "membrane anchor  
acids 10-29)" hydrophobic amino acid stretch in GntI

OTHER INFORMATION: /function  
acids 10-29)"  
"hydrophobic amino acid stretch in GnTI"  
of a type II

```

/product= "hydrophobic amino acid stretch in Gln
"membrane anchor of a type II

```

```
/product= "hydrophobic membrane anchor of a type 1  
/standard_name= "golgi protein"
```

FEATURE:  
(A) NAME/KEY: misc\_feature

(A) NAME/KEY: MISS-  
(B) LOCATION: 1..14  
INFORMAT

(A) NAME/KEY: misc\_feature  
(B) LOCATION:1..14  
(D) OTHER INFORMATION:/function="use for cloning the  
our library in Lambda ZAPII"  
"adapter"

OTHER INFORMATION: /function- as  
cDNA library in Lambda ZAPII"  
.../NotI-cDNA ad

```
cdna library in Lambda
/product= "EcoRI/NotI-cDNA adapter"
```

```
/product=
/number= 1
```

FEATURE:  
(A) NAME/KEY: misc\_feature

(A) NAME/KEY: MISC-100-  
(B) LOCATION: 1724..1737  
INFORMATION: /

(D) OTHER INFORMATION: /product= "EcoRI/NotI-cDNA adapt  
/number= 2

/number= 2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:	60
GAATTCGCGG CCGCCATTGA CTTGATCCTA ACTGAACAGG CAAAGTAAAT CCAGCGATGA	120
TCGCTTTCTC CTAAAGCCTT CAATCGAATT	180

GAATTCGCGG CCGCCATTGA CTTGATCCTA ACTGAACAGG CTTTGGTCTG TGTGGTGGT  
AACACTCATA ACTGAACACT GAGAGACTAT TCGCTTTCTC CTAAAGCCTT CAATCGAATT  
TTT TGC TGT GAT TTC CGG TAC CTC CTC

AAACATCATA ACTGAACACT GAGAGACTAT TCGATTTA 168  
CGCACG ATG AGA GGG AAC AAG TTT TGC TGT GAT TTC CGG TAC CTC CTC  
Met Arg Gly Asn Lys Phe Cys Cys Asp Phe Arg Tyr Leu Leu 460  
450 455 21

Met Arg Gly 450  
ATC TTG GCT GCT GTC GCC TTC ATC TAC ACA CAG ATG CGG CTT TTT GCG  
Ile Leu Ala Ala Val Ala Phe Ile Tyr Thr Gln Met Arg Leu Phe Ala 475  
465  
ATT GAA GCA GAA 264

Ile Leu Ala Ala Val  
465  
ACA CAG TCA GAA TAT GCA GAT CGC CTT GCT GCT GCA ATT GAA GCA GAA  
Thr Gln Ser Glu Tyr Ala Asp Arg Leu Ala Ala Ala Ile Glu Ala Glu  
480  
485  
490  
312

[illegible]

Asn His Cys Thr Ser 500 360  
495  
CAG CAA GGA AGA ATA GTT GCT CTT GAA GAA CAA ATG AAG CGT CAG GAC  
Gln Gln Gly Arg Ile Val Ala Leu Glu Glu Gln Met Lys Arg Gln Asp  
510 515 520 525 400  
CCT GCT GCT CAG GAT CTT GAA AGT AAG  
Leu Ser Lys

Gln Gln Gly Arg 515  
510  
CAG GAG TGC CGA CAA TTA AGG GCT CTT GTT CAG GAT CTT GAA AGT AAG  
Gln Glu Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys 540  
530  
535  
540  
545

Gln Glu Cys Arg Gln 530 535 456  
GGC ATA AAA AAG TTG ATC GGA AAT GTA CAG ATG CCA GTG GCT GCT GTA  
Gly Ile Lys Lys Leu Ile Gly Asn Val Gln Met Pro Val Ala Ala Val 555  
545 550 504

Gly Ile Lys Lys 545 550 504  
GTT GTT ATG GCT TGC AAT CGG GCT GAT TAC CTG GAA AAG ACT ATT AAA  
Val Val Met Ala Cys Asn Arg Ala Asp Tyr Leu Glu Lys Thr Ile Lys 570 552

[illegible]

Ser Ile Leu Lys 580  
 575  
 ATA TCC CAG GAT GGA TCA CAT CCT GAT GTC AGG AAG CTT GCT TTG AGC  
 Ile Ser Gln Asp Gly Ser His Pro Asp Val Arg Lys Leu Ala Leu Ser  
 590 595 600  
 648

590  
 TAT GAT CAG CTG ACG TAT ATG CAG CAC TTG GAT TTT GAA CCT GTG CAT  
 Tyr Asp Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val His  
 610  
 615  
 620  
 690

Tyr Asp Gln Leu 610  
 ACT GAA AGA CCA GGG GAG CTG ATT GCA TAC TAC AAA ATT GCA CGT CAT 696  
 Thr Glu Arg Pro Gly Glu Leu Ile Ala Tyr Tyr Lys Ile Ala Arg His 635  
 625  
 630  
 640  
 650  
 660  
 670  
 680  
 690  
 700  
 710  
 720  
 730  
 740  
 750  
 760  
 770  
 780  
 790  
 800  
 810  
 820  
 830  
 840  
 850  
 860  
 870  
 880  
 890  
 900  
 910  
 920  
 930  
 940  
 950  
 960  
 970  
 980  
 990  
 1000

Thr Glu Arg Phe Gly 630  
625  
TAC AAG TGG GCA TTG GAT CAG CTG TTT TAC AAG CAT AAT TTT AGC CGT 744  
Tyr Lys Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Asn Phe Ser Arg  
640 645 650  
GCG CGT GAT TTT TTT GAC 792

Tyr Lys Trp Ala Leu Met 645  
 640  
 GTT ATC ATA CTA GAA GAT GAT ATG GAA ATT GCC CCT GAT TTT TTT GAC 792  
 Val Ile Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp  
 655  
 660  
 840

Val Ile Ile Leu Glu Asp 660  
655  
TTT TTT GAG GCT GGA GCT ACT CTT CTT GAC AGA GAC AAG TCG ATT ATG 840



Phe Phe Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met  
 670 675 680 888  
 GCT ATT TCT TCT TGG AAT GAC AAT GGA CAA ATG CAG TTT GTC CAA GAT  
 Ala Ile Ser Ser Trp Asn Asp Asn Gly Gln Met Gln Phe Val Gln Asp  
 690 700  
 CCT TAT GCT CTT TAC CGC TCA GAT TTT TTT CCC GGT CTT GGA TGG ATG  
 Pro Tyr Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met  
 705 710 715 936  
 CTT TCA AAA TCT ACT TGG GAC GAA TTA TCT CCA AAG TGG CCA AAG GCT  
 Leu Ser Lys Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala  
 720 725 730 984  
 TAC TGG GAC GAC TGG CTA AGA CTC AAA GAG AAT CAC AGA GGT CGA CAA  
 Tyr Trp Asp Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln  
 735 740 745 1032  
 TTT ATT CGC CCA GAA GTT TGC AGA ACA TAT AAT TTT GGT GAG CAT GGT  
 Phe Ile Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly  
 750 755 760 765 1080  
 TCT AGT TTG GGG CAG TTT TTC AAG CAG TAT CTT GAG CCA ATT AAA CTA  
 Ser Ser Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu  
 770 775 780 1128  
 AAT GAT GTC CAG GTT GAT TGG AAG TCA ATG GAC CTT AGT TAC CTT TTG  
 Asn Asp Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu  
 785 790 795 1176  
 GAG GAC AAT TAC GTG AAA CAC TTT GGT GAC TTG GTT AAA AAG GCT AAG  
 Glu Asp Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys  
 800 805 810 1224  
 CCC ATC CAT GGA GCT GAT GCT GTC TTG AAA GCA TTT AAC ATA GAT GGT  
 Pro Ile His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly  
 815 820 825 1272  
 GAT GTG CGT ATT CAG TAC AGA GAT CAA CTA GAC TTT GAA AAT ATC GCA  
 Asp Val Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala  
 830 835 840 845 1320  
 CGG CAA TTT GGC ATT TTT GAA GAA TGG AAG GAT GGT GTA CCA CGT GCA  
 Arg Gln Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala  
 850 855 860 1368  
 GCA TAT AAA GGA ATA GTA GTT TTC CGG TAC CAA ACG TCC AGA CGT GTA  
 Ala Tyr Lys Gly Ile Val Val Phe Arg Tyr Gln Thr Ser Arg Arg Val  
 865 870 875 1416  
 TTC CTT GTT GGC CAT GAT TCG CTT CAA CAA CTC GGA ATT GAA GAT ACT  
 Phe Leu Val Gly His Asp Ser Leu Gln Gln Leu Gly Ile Glu Asp Thr  
 880 885 890 1464  
 TAA CAAAGATATG ATTGCAGGAG CCCGGGCAAA ATTTTGTACT TATTGGGTAG  
 \* 1517  
 GATGCATCGA GCTGACACTA AACCATGATT TTACCAGTTA CATAACAGT TTTAATGTTA 1577  
 TACGGAGGAG CTCACGTGTT TAGTGTTGAA GGGATATCGG CTTCTTAGTA TTGGATGAAT 1637

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:  
(ii) LENGTH: 447 amino acid

- SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 447 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION:

(ii) MOLECULE TYPE: protein  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

(11) MOLECULAR WEIGHT: 10,000  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1

Met Arg Gly Asn Lys Phe Cys Cys Asp Phe Arg Tyr Leu Leu Ile Leu 15  
1 5 10

Ala Ala Val Ala Phe Ile Tyr Thr Gln Met Arg Leu Phe Ala Thr Gln 30  
20 25

Ser Glu Tyr Ala Asp Arg Leu Ala Ala Ile Glu Ala Glu Asn His 45  
35 40

Cys Thr Ser Gln Thr Arg Leu Leu Ile Asp Gln Ile Ser Leu Gln Gln 60  
50 55

Gly Arg Ile Val Ala Leu Glu Glu Gln Met Lys Arg Gln Asp Gln Glu 80  
65 70 75

Cys Arg Gln Leu Arg Ala Leu Val Gln Asp Leu Glu Ser Lys Gly Ile 95  
85 90

Lys Lys Leu Ile Gly Asn Val Gln Met Pro Val Ala Ala Val Val Val 110  
100 105

Met Ala Cys Asn Arg Ala Asp Tyr Leu Glu Lys Thr Ile Lys Ser Ile 125  
115 120

Leu Lys Tyr Gln Ile Ser Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser 140  
130 135

Gln Asp Gly Ser His Pro Asp Val Arg Lys Leu Ala Leu Ser Tyr Asp 160  
145 150 155

Gln Leu Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val His Thr Glu 175  
165 170

Arg Pro Gly Glu Leu Ile Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys 190  
180 185

Trp Ala Leu Asp Gln Leu Phe Tyr Lys His Asn Phe Ser Arg Val Ile 205  
195 200

Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp Phe Phe 220  
210 215

Glu Ala Gly Ala Thr Leu Leu Asp Arg Asp Lys Ser Ile Met Ala Ile 240  
225 230 235

Ser Ser Trp Asn Asp Asn Gly Gln Met Gln Phe Val Gln Asp Pro Tyr 255  
245 250

Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Ser  
260 265 270

Lys Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp  
275 280 285

Asp Asp Trp Leu Arg Leu Lys Glu Asn His Arg Gly Arg Gln Phe Ile  
290 295 300

Arg Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser  
305 310 315 320

Leu Gly Gln Phe Phe Lys Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp  
325 330 335

Val Gln Val Asp Trp Lys Ser Met Asp Leu Ser Tyr Leu Leu Glu Asp  
340 345 350

Asn Tyr Val Lys His Phe Gly Asp Leu Val Lys Lys Ala Lys Pro Ile  
355 360 365

His Gly Ala Asp Ala Val Leu Lys Ala Phe Asn Ile Asp Gly Asp Val  
370 375 380

Arg Ile Gln Tyr Arg Asp Gln Leu Asp Phe Glu Asn Ile Ala Arg Gln  
385 390 395 400

Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Ala Ala Tyr  
405 410 415

Lys Gly Ile Val Val Phe Arg Tyr Gln Thr Ser Arg Arg Val Phe Leu  
420 425 430

Val Gly His Asp Ser Leu Gln Gln Leu Gly Ile Glu Asp Thr \*  
435 440 445

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1854 base pairs
  - (B) TYPE: Nucleotide
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: No
- (iv) ANTI-SENSE: No
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Arabidopsis thaliana
  - (B) STRAIN: Columbia
  - (D) DEVELOPMENTAL STAGE: Mature plants
  - (F) TISSUE TYPE: All tissues
- (vii) IMMEDIATE SOURCE:
  - (A) LIBRARY: Lambda Uni-ZAP (EcoRI/XhoI) and  
Lambda ACT (XhoI)
  - (B) CLONE: pBSK-Ara-GntI-full #8
- (ix) FEATURE:

- (ix) FEATURE:

- (ix) FEATURE:

- (ix) FEATURE:

- (ix) FEATURE:

- (ix) FEATURE:

- (ix) FEATURE:

- ```

FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION:1849..1854
(D) OTHER INFORMATION:/product= "XhoI-cDNA-Adaptor"
    /number= 2

```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

|            |            |                         |                     |            |             |     |
|------------|------------|-------------------------|---------------------|------------|-------------|-----|
| CTCGAGGCCA | CGAAGGCCAC | CGTTTTTGTT              | ATAACGAACG          | ACACCGTTTC | AAACAAC TTC | 60  |
| CTTATTAGCT | AGCTCCCTCC | CGGCGGCAAA              | CACCAGAAGA          | TCCACCGCTT | TTGATCTGGT  | 120 |
|            |            | GGC AGG ATC TCG TGT GAC | TTG AGA TTT CTT CTC |            |             | 170 |

Met Ala Arg 5  
1  
ATC CCG GCA GCT TTC ATG TTC ATC TAC ATC CAG ATG AGG CTT TTC CAG 218  
Ile Pro Ala Ala Phe Met Phe Ile Tyr Ile Gln Met Arg Leu Phe Gln 25  
15  
20  
26

AAC CAT TGC ACT AGT CAA ATG CGA GGC CTC ATA GAT GAA GTT AGC ATC  
 Asn His Cys Thr Ser Gln Met Arg Gly Leu Ile Asp Glu Val Ser Ile 60

Lys Gln Ser Arg Ile Val  
 65  
 GAA GAA CTT GTG CAG CTT AAG GAT CTA ATC CAG ACG TTT GAA AAA AAA  
 Glu Glu Leu Val Gln Leu Lys Asp Leu Ile Gln Thr Phe Glu Lys Lys  
 80  
 85  
 90  
 95  
 100  
 105  
 110  
 115  
 120  
 125  
 130  
 135  
 140  
 145  
 150  
 155  
 160  
 165  
 170  
 175  
 180  
 185  
 190  
 195  
 200  
 205  
 210  
 215  
 220  
 225  
 230  
 235  
 240  
 245  
 250  
 255  
 260  
 265  
 270  
 275  
 280  
 285  
 290  
 295  
 300  
 305  
 310  
 315  
 320  
 325  
 330  
 335  
 340  
 345  
 350  
 355  
 360  
 365  
 370  
 375  
 380  
 385  
 390  
 395  
 400  
 405  
 410  
 415  
 420  
 425  
 430  
 435  
 440  
 445  
 450  
 455  
 460  
 465  
 470  
 475  
 480  
 485  
 490  
 495  
 500  
 505  
 510  
 515  
 520  
 525  
 530  
 535  
 540  
 545  
 550  
 555  
 560  
 565  
 570  
 575  
 580  
 585  
 590  
 595  
 600  
 605  
 610  
 615  
 620  
 625  
 630  
 635  
 640  
 645  
 650  
 655  
 660  
 665  
 670  
 675  
 680  
 685  
 690  
 695  
 700  
 705  
 710  
 715  
 720  
 725  
 730  
 735  
 740  
 745  
 750  
 755  
 760  
 765  
 770  
 775  
 780  
 785  
 790  
 795  
 800  
 805  
 810  
 815  
 820  
 825  
 830  
 835  
 840  
 845  
 850  
 855  
 860  
 865  
 870  
 875  
 880  
 885  
 890  
 895  
 900  
 905  
 910  
 915  
 920  
 925  
 930  
 935  
 940  
 945  
 950  
 955  
 960  
 965  
 970  
 975  
 980  
 985  
 990  
 995  
 1000

100  
 Gly Ile Ala Lys Leu Thr Ser  
 95  
 GTG GTT ATG GCC TGC AGT CGT GCA GAC TAT CTT GAA AGG ACT GTT AAA  
 Val Val Met Ala Cys Ser Arg Ala Asp Tyr Leu Glu Arg Thr Val Lys  
 110  
 115  
 55  
 CCG GTT GCT TCA AAA TAT CCT CTA TTT  
 Phe Pro Leu Phe

Ser Val Leu Thr Tyr 130  
 125  
 ATA TCT CAG GAT GGA TCT GAT CAA GCT GTC AAG AGC AAG TCA TTG AGC  
 Ile Ser Gln Asp Gly Ser Asp Gln Ala Val Lys Ser Lys Ser Leu Ser 155  
 145  
 TTT GAA CCA GTG GTC  
 150  
 65

ACT GAA AGG CCT GGT GAA CTG ACT GCG TAC TAC AAG ATT GCA CGT CAC  
Thr Glu Arg Pro Gly Glu Leu Thr Ala Tyr Tyr Lys Ile Ala Arg His

Tyr Lys Trp Ala Leu Asp 195  
 190  
 GTG ATT ATA CTA GAA GAC GAT ATG GAA ATT GCT CCA GAC TTC TTT GAT  
 Val Ile Ile Leu Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp 220  
 205 210 215

|                                                                                                                                                       |              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| TAC TTT GAG GCT GCA GCT AGT CTC ATG GAT AGG GAT AAA ACC ATT ATG<br>Tyr Phe Glu Ala Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met<br>225 230 235     | 842          |
| GCT GCT TCA TCA TGG AAT GAT AAT GGA CAG AAG CAG TTT GTG CAT GAT<br>Ala Ala Ser Ser Trp Asn Asp Asn Gly Gln Lys Gln Phe Val His Asp<br>240 245 250     | 890          |
| CCC TAT GCG CTA TAC CGA TCA GAT TTT TTT CCT GGC CTT GGG TGG ATG<br>Pro Tyr Ala Leu Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met<br>255 260 265     | 938          |
| CTC AAG AGA TCG ACT TGG GAT GAG TTA TCA CCA AAG TGG CCA AAG GCT<br>Leu Lys Arg Ser Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala<br>270 275 280     | 986          |
| TAC TGG GAT GAT TGG CTG AGA CTA AAG GAA AAC CAT AAA GGC CGC CAA<br>Tyr Trp Asp Asp Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln<br>285 290 295 300 | 1034         |
| TTC ATT GCA CCG GAA GTC TGT AGA ACA TAC AAT TTT GGT GAA CAT GGG<br>Phe Ile Ala Pro Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly<br>305 310 315     | 1082         |
| TCT AGT TTG GGA CAG TTT TTC AGT CAG TAT CTG GAA CCT ATA AAG CTA<br>Ser Ser Leu Gly Gln Phe Phe Ser Gln Tyr Leu Glu Pro Ile Lys Leu<br>320 325 330     | 1130         |
| AAC GAT GTG ACG GTT GAC TGG AAA GCA AAG GAC CTG GGA TAC CTG ACA<br>Asn Asp Val Thr Val Asp Trp Lys Ala Lys Asp Leu Gly Tyr Leu Thr<br>335 340 345     | 1178         |
| GAG GGA AAC TAT ACC AAG TAC TTT TCT GGC TTA GTG AGA CAA GCA CGA<br>Glu Gly Asn Tyr Thr Lys Tyr Phe Ser Gly Leu Val Arg Gln Ala Arg<br>350 355 360     | 1226         |
| CCA ATT CAA GGT TCT GAC CTT GTC TTA AAG GCT CAA AAC ATA AAG GAT<br>Pro Ile Gln Gly Ser Asp Leu Val Leu Lys Ala Gln Asn Ile Lys Asp<br>365 370 375 380 | 1274         |
| GAT GAT CGT ATC CGG TAT AAA GAC CAA GTA GAG TTT GAA CGC ATT GCA<br>Asp Asp Arg Ile Arg Tyr Lys Asp Gln Val Glu Phe Glu Arg Ile Ala<br>385 390 395     | 1322         |
| GGG GAA TTT GGT ATA TTT GAA GAA TGG AAG GAT GGT GTG CCA CGA ACA<br>Gly Glu Phe Gly Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Thr<br>400 405 410     | 1370         |
| GCA TAT AAA GGA GTA GTG GTG TTT CGA ATC CAG ACA ACA AGA CGT GTA<br>Ala Tyr Lys Gly Val Val Phe Arg Ile Gln Thr Thr Arg Arg Val<br>415 420 425         | 1418         |
| TTC CTG GTT GGG CCA GAT TCT GTA ATG CAG CTT GGA ATT CGA AAT TCC<br>Phe Leu Val Gly Pro Asp Ser Val Met Gln Leu Gly Ile Arg Asn Ser<br>430 435 440     | 1466         |
| TGA TGCAAAACAT ATGAAAGGAA AAGAAGATTT TGGACCGCAT GCAGCCTCCT<br>*<br>445                                                                                | 1519         |
| TCTAGCAGCT GTTAGGTTGT ATTGTTATTT ATGGATGAGT TTGTAGAGCG GTGGGGTTAA<br>CTTTAACAGC AAGGAAGCTC TGGTGACCAG GCTGATTGGC TTAGAAGTTA TGGGAACCCC                | 1579<br>1639 |

TTGAAAGGGT CAGGGTTAAA TATATTTTCAG TTGTTTTTATT AGTGATTATC TTGTGGGTAA 1699  
 CTTATACGAA TGCAAATCAT TCTATGCAGT TTTTCTTCGT CCCACTTGTT TTGGCTTCTC 1759  
 TATTGCTAGT GTACATATCT CTTCAAACAT GTACTAAATA ATGCGTGTG CTTCAAAGAA 1819  
 GTAACCTTTTA TTAAAAAAA AAAAAAAC TCGAG 1854

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 445 amino acids  
 (B) TYPE: Amino acid  
 (D) TOPOLOGY: Linear

- (ii) MOLECULE TYPE: Protein  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Ala Arg Ile Ser Cys Asp Leu Arg Phe Leu Leu Ile Pro Ala Ala 15  
 1 5 10  
 Phe Met Phe Ile Tyr Ile Gln Met Arg Leu Phe Gln Thr Gln Ser Gln 30  
 20 25 30  
 Tyr Ala Asp Arg Leu Ser Ser Ala Ile Glu Ser Glu Asn His Cys Thr 45  
 35 40 45  
 Ser Gln Met Arg Gly Leu Ile Asp Glu Val Ser Ile Lys Gln Ser Arg 60  
 50 55 60  
 Ile Val Ala Leu Glu Asp Met Lys Asn Arg Gln Asp Glu Glu Leu Val 80  
 65 70 75  
 Gln Leu Lys Asp Leu Ile Gln Thr Phe Glu Lys Lys Gly Ile Ala Lys 95  
 85 90 95  
 Leu Thr Gln Gly Gly Gln Met Pro Val Ala Ala Val Val Val Met Ala 110  
 100 105 110  
 Cys Ser Arg Ala Asp Tyr Leu Glu Arg Thr Val Lys Ser Val Leu Thr 125  
 115 120 125  
 Tyr Gln Thr Pro Val Ala Ser Lys Tyr Pro Leu Phe Ile Ser Gln Asp 140  
 130 135 140  
 Gly Ser Asp Gln Ala Val Lys Ser Lys Ser Leu Ser Tyr Asn Gln Leu 160  
 145 150 155  
 Thr Tyr Met Gln His Leu Asp Phe Glu Pro Val Val Thr Glu Arg Pro 175  
 165 170 175  
 Gly Glu Leu Thr Ala Tyr Tyr Lys Ile Ala Arg His Tyr Lys Trp Ala 190  
 180 185 190  
 Leu Asp Gln Leu Phe Tyr Lys His Lys Phe Ser Arg Val Ile Ile Leu 205  
 195 200 205  
 Glu Asp Asp Met Glu Ile Ala Pro Asp Phe Phe Asp Tyr Phe Glu Ala 220  
 210 215 220  
 Ala Ala Ser Leu Met Asp Arg Asp Lys Thr Ile Met Ala Ala Ser Ser

000000-000000

230 235 240

225

Trp Asn Asp Asn Gly Gln Lys Gln Phe Val His Asp Pro Tyr Ala Leu  
245 250 255

Tyr Arg Ser Asp Phe Phe Pro Gly Leu Gly Trp Met Leu Lys Arg Ser  
260 265 270

Thr Trp Asp Glu Leu Ser Pro Lys Trp Pro Lys Ala Tyr Trp Asp Asp  
275 280 285

Trp Leu Arg Leu Lys Glu Asn His Lys Gly Arg Gln Phe Ile Ala Pro  
290 295 300

Glu Val Cys Arg Thr Tyr Asn Phe Gly Glu His Gly Ser Ser Leu Gly  
305 310 315 320

Gln Phe Phe Ser Gln Tyr Leu Glu Pro Ile Lys Leu Asn Asp Val Thr  
325 330 335

Val Asp Trp Lys Ala Lys Asp Leu Gly Tyr Leu Thr Glu Gly Asn Tyr  
340 345 350

Thr Lys Tyr Phe Ser Gly Leu Val Arg Gln Ala Arg Pro Ile Gln Gly  
355 360 365

Ser Asp Leu Val Leu Lys Ala Gln Asn Ile Lys Asp Asp Arg Ile  
370 375 380

Arg Tyr Lys Asp Gln Val Glu Phe Glu Arg Ile Ala Gly Glu Phe Gly  
385 390 395 400

Ile Phe Glu Glu Trp Lys Asp Gly Val Pro Arg Thr Ala Tyr Lys Gly  
405 410 415

Val Val Val Phe Arg Ile Gln Thr Thr Arg Arg Val Phe Leu Val Gly  
420 425 430

Pro Asp Ser Val Met Gln Leu Gly Ile Arg Asn Ser \*  
435 440 445